En route to waveform inversion by MBTT formulation : optimization for one shot and gradient for multiple shots

Guy Chavent, François Clément, Susana Gomez

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

En route to waveform inversion by MBTT formulation : optimization for one shot and gradient for multiple shots

Guy CHAVENT, François CLÉMENT,
Susana GOMEZ

N° 2150
Janvier 1994

PROGRAMME 6

Calcul scientifique,
modélisation et
logiciels numériques

En route to waveform inversion by MBTT formulation: optimization for one shot and gradient for multiple shots*

Vers l'inversion sismique par une reformulation en temps : optimisation pour un tir et gradient pour tirs multiples

Guy CHAVENT[†‡]
François CLÉMENT[‡]
Susana GÓMEZ[‡]

*Work performed as part of the IFP-INRIA Ψ Consortium Project.
[†]CEREMADE, Université de Paris-Dauphine, pl. du M[al] De Lattre De Tassigny, F-75775 Paris Cédex 16.
[‡]INRIA-Rocquencourt, Domaine de Voluceau, B.P. 105, F-78153 Le Chesnay Cédex.

ABSTRACT

This work is devoted to automatic determination of 2D background slowness from full waveform data. We have designed, in earlier works, an original approach, the Migration-Based Travel Time (MBTT) reformulation of the waveform acoustic inversion problem and we give here its present status of development.

After a brief recall of the principle of this approach, we address three important issues: first, we prove the legitimacy of the MBTT change of reflectivity unknowns, then, we illustrate the ability of our MBTT criterion to be minimized with respect to background slowness by local techniques in the single shot case, and finally, we show that, in the case of multiple shots, the first gradient calculation is very promising for future minimization.

We also show a validation of our gradient and we describe a Limited Memory Quasi-Newton minimization method.

Key words: inverse problem, parameter estimation, seismic inversion, background velocities determination, travel-time formulation, migration, multiscale analysis, optimization, limited memory BFGS method.

RÉSUMÉ

Ces travaux sont consacrés à la détermination automatique de modèles de lenteur bidimensionnels à partir de données sismiques complètes. Nous avons développé, dans des travaux antérieurs, une approche originale, la reformulation en temps de parcours par migration (MBTT) du problème inverse sismique et nous en montrons ici l'état d'avancement.

Après un bref rappel du principe de cette approche, nous abordons trois points importants : nous montrons d'abord la légitimité du changement d'inconnue réflectivité par MBTT, puis nous illustrons la capacité de notre critère MBTT à être minimisé par rapport au modèle de lenteur par des techniques locales dans le cas d'un seul tir, et enfin nous montrons que, dans le cas de tirs multiples, le premier pas de gradient est déjà très prometteur pour l'optimisation future.

Nous montrons aussi une validation du gradient et nous décrivons une méthode de minimisation de type Quasi-Newton à mémoire limitée.

Mots clefs : problème inverse, estimation de paramètres, inversion sismique, détermination de modeles de vitesses, reformulation en temps de parcours, migration, analyse multi-échelle, optimisation, méthode BFGS à mémoire limitée.

1 INTRODUCTION

The geophysical problem of the determination of the Earth material properties from seismic surface data is commonly mathematically set as an inverse problem. The direct problem modelizes the real field experiment and the inverse problem, usually set as a minimization problem, consists in finding the Earth material properties that reproduce at best the seismic data through the direct model.

The difficulty is that local optimization methods are inefficient when applied to this minimization problem as it presents many local minima, which are meaningless, and that global optimization methods are far too expensive, due to the huge amount of data and unknowns to handle.

Considering that global methods were not for today, we could only modify the function to minimize and we have designed (Clément, 1991) an original approach by introducing a new "time" reflectivity unknown: the Migration-Based Travel Time (MBTT) reformulation. Extended in Clément and Chavent (1992) to the multishot case, it is related to the works of Chavent and Jacewitz (1990), Symes and Kern (1992) and Bunks (1993).

First, we briefly recall, in section 2, the MBTT reformulation of the seismic inverse problem, then we prove theoretically and numerically the legitimacy of this reformulation (section 3). In section 4, we present the numerical results of the multiscale minimization with respect to the background slowness in the single shot case. And finally, the case of multiple shots is presented in section 5 and we show the gradient of the modified data misfit function. In appendix, we present a test used to check the computation of the gradient and we describe the Limited Memory BFGS optimization method.

2 THE MBTT APPROACH TO FULL WAVEFORM INVERSION

Here, we shortly recall the principle of the MBTT approach to full waveform inversion, and refer to Clément and Chavent (1992) for a detailed description. Given material properties ν and σ (slowness and impedance as functions of the horizontal distance x and the depth z), we denote by $c = \varphi(\nu, \sigma)$ the corresponding (collection of, in case of more than one shot) synthetic seismogram(s). Thus φ is the forward modelling operator (wave equation), and the usual least squares data misfit function for a given collection d of seismic data is

$$\mathcal{J}(\nu, \sigma) = \frac{1}{2}\|d - \varphi(\nu, \sigma)\|^2 \tag{2.1}$$

where the norm $\|.\|$ on the data space takes into account muting and correction for geometrical spreading.

As the information on the low frequencies of the slowness model is available only through the kinematics of the reflected energy, one first performs a smooth/rough decomposition of ν and σ

$$\nu = \nu_s + \nu_r \tag{2.2}$$

$$\sigma = \sigma_s + \sigma_r, \tag{2.3}$$

in such a way (Clément and Chavent, 1992, 1993), that ν_s describes mostly the kinematics (propagator unknown) and $r = (\nu_r, \sigma_r)$ describes mostly the reflectivity of the medium

(reflectivity unknown). The smooth part σ_s of the impedance has little influence on the seismic sections, and is supposed known. One can now obviously replace the search for the *material unknowns* (ν, σ) by the search for the *intermediate unknowns* (ν_s, r). When expressed as a function of the intermediate unknowns, the synthetic c is denoted by $\mathcal{I}\varphi(\nu_s, r)$ and the data misfit function by $\mathcal{I}\mathcal{J}(\nu_s, r) = \frac{1}{2}\|d - \mathcal{I}\varphi(\nu_s, r)\|^2$. The error function $\mathcal{I}\mathcal{J}$ is known to be almost quadratical in the reflectivity unknown r, but extremely non convex in the propagator unknown ν_s, because of the phase shifts induced in $\mathcal{I}\varphi(\nu_s, r)$ by changes in ν_s. These phase shifts generate many local minima of $\mathcal{I}\mathcal{J}$, thus prohibiting its minimization by local optimization techniques.

The MBTT solution to this consists in getting rid of these phase shifts by replacing the search for a "depth" reflectivity r, function of (x, z), by the search for a "time" reflectivity s, function of (x, t). In the one-dimensional wave equation, this is basically done by replacing the depth variable z by the travel time variable τ. In two- (or three- !) dimensional wave equation, this simple change of variables is no more possible, as the travel time depends on the geometry of sources and receivers and because there may be more than one ray path from one point to the other. But, when the geometry is fixed, as in a given shot, the migration allows to transform an (x, t) time section into an (x, z) depth section. So the MBTT approach replaces the search for $r = (\nu_r, \sigma_r)$ by the search for NSHOT time sections $s_1, \ldots, s_{\text{NSHOT}}$ such that

$$r = \sum_{n=1}^{\text{NSHOT}} r_n \tag{2.4}$$

with

$$r_n = \mathcal{M}_n(\nu_s) s_n, \qquad \text{for all } n = 1, \ldots, \text{NSHOT} \tag{2.5}$$

where $\mathcal{M}_n(\nu_s)$ is a migration operator for the nth shot with migration slowness ν_s. We describe here only the simplest "stacking" version of MBTT, which is the one for which numerical experiments are shown below; another version without stack but with a semblance penalty term is also possible (Clément and Chavent, 1992).

When expressed as a function of the *modified unknowns* $(\nu_s, s = (s_1, \ldots, s_{\text{NSHOT}}))$, the synthetic seismogram c is denoted by $\mathcal{M}\varphi(\nu_s, s)$, and the data misfit function by $\mathcal{M}\mathcal{J}(\nu_s, s) = \frac{1}{2}\|d - \mathcal{M}\varphi(\nu_s, s)\|^2$.

The MBTT approach consists thus in replacing the resolution of the problem

$$(\mathcal{P}) \qquad \inf_{(\nu,\sigma)} \mathcal{J}(\nu, \sigma) \quad \text{or} \quad (\mathcal{I}\mathcal{P}) \qquad \inf_{(\nu_s, r)} \mathcal{I}\mathcal{J}(\nu_s, r) \tag{2.6}$$

by that of the modified problem

$$(\mathcal{M}\mathcal{P}) \qquad \inf_{(\nu_s, s)} \mathcal{M}\mathcal{J}(\nu_s, s). \tag{2.7}$$

The questions that arise naturally are:

- is it legitimate to replace $(\mathcal{P})$ by $(\mathcal{M}\mathcal{P})$? In other terms, do we not underparametrize the reflectivity by replacing the depth reflectivity r by the time reflectivity s ?
- Does it do any good to the phase shift problem ? Our measure of satisfaction will be to check that we can minimize $\mathcal{M}\mathcal{J}$ with respect to ν_s by local methods

without encountering local minima—*i.e.* that do not approximate data with enough accuracy—even in the case of a single shot. Of course we do not expect, in that case, to find a unique global minimum for the propagator ν_s, but we want to check that we can bring the value of the error function down to its global minimum value—*i.e.* fully explain the single shot gather—by local optimization methods.

- In the case of multiple shots, is it possible to extract the kinematic information hidden in the data through the minimization of $\mathcal{MJ}$ with respect to ν_s by a local method, for fixed and easy-to-obtain values of the time reflectivity ?

We address now these three questions.

3 LEGITIMACY OF THE MBTT CHANGE OF UNKNOWN

Let $(\bar{\nu}_s, \bar{r})$ be the true intermediate parameters which have generated the noiseless data $d = (d_1, \ldots, d_{\text{NSHOT}})$. Then, as r is parametrized by s, one obviously has

$$0 = \inf_{(\nu_s, r)} \mathcal{IJ}(\nu_s, r) \leq \inf_{(\nu_s, s)} \mathcal{MJ}(\nu_s, s). \tag{3.1}$$

The change of unknown will be legitimate if the inequality is in fact an equality, which would be obviously the case if there exists an $\bar{s} = (\bar{s}_1, \ldots, \bar{s}_{\text{NSHOT}})$ such that

$$\bar{r} = \sum_{n=1}^{\text{NSHOT}} \mathcal{M}_n(\bar{\nu}_s)\bar{s}_n. \tag{3.2}$$

Of course this is impossible in general, as the migration operators $\mathcal{M}_n(\bar{\nu}_s)$ for each shot are not onto: for example, it is impossible to satisfy the above equality if $\bar{r}$ contains a reflector which is illuminated by none of the shots !

Hence requiring that (3.2) holds is demanding too much. We shall see that we can have the equality in (3.1) even if (3.2) is not satisfied.

An important feature for this is to choose the migration operator for the shot n as

$$\mathcal{M}_n(\nu_s) = H_n \mathcal{I}\varphi_{n,r}(\nu_s, 0)^T \tag{3.3}$$

where $\mathcal{I}\varphi_{n,r}(\nu_s, 0)$ is the forward linearized map of the nth shot and H_n is a symmetric positive definite operator, chosen in such a way that $\mathcal{M}_n(\nu_s)$ is a "true amplitude" migration. The above choice for $\mathcal{M}_n(\nu_s)$ insures that the migrated residual $\mathcal{M}_n(\nu_s)(d_n - \mathcal{I}\varphi_n(\nu_s, 0)) \simeq \mathcal{M}_n(\nu_s)d_n$ (remember that $\mathcal{I}\varphi_n(\nu_s, 0) \simeq 0$ by definition of the smooth/rough decomposition of ν) is a descent direction for the error function $\mathcal{IJ}_n$ of the nth shot at $r = 0$ (Chavent, 1993).

3.1 CASE OF A SINGLE SHOT

First, we consider the case of a single shot—and hence drop the index n. We give three reasons for which equality should hold in (3.1).

3.1.1 THE RESIDUAL: A GOOD FIRST GUESS OF $\bar{s}$

As it was shown in Clément and Chavent (1992), the residual

$$s^0 = d - \mathcal{I}\varphi(\bar{\nu}_s, 0) \simeq d \tag{3.4}$$

gives a good first guess of $\bar{s}$, as the corresponding resimulated seismogram $\mathcal{M}\varphi(\bar{\nu}_s, s^0)$ is quite close to the data d, so that $\mathcal{M}\mathcal{J}(\bar{\nu}_s, s^0)$ is already small.

3.1.2 LINEARIZED FORWARD MODEL: THEORETICAL RESULT

The replacement of r by s is fully mathematically justified when a linearized forward model is used. Let us linearize the forward model $\mathcal{I}\varphi(\nu_s, r)$ with respect to reflectivity at $\nu_s = \bar{\nu}_s$ and zero background reflectivity, and replace r by the scaled reflectivity

$$\tilde{r} = H^{-\frac{1}{2}} r. \tag{3.5}$$

When this scaled reflectivity is used, the migration operator (3.3) reduces to

$$\widetilde{\mathcal{M}}(\bar{\nu}_s) = \mathcal{I}\varphi_{,\tilde{r}}(\bar{\nu}_s, 0)^T \tag{3.6}$$

and the linearized error functions write

$$\mathcal{L}\mathcal{I}\mathcal{J}(\bar{\nu}_s, \tilde{r}) = \frac{1}{2}\|\mathcal{L}d - \mathcal{I}\varphi_{,\tilde{r}}(\bar{\nu}_s, 0)\tilde{r}\|^2 \tag{3.7}$$

$$\mathcal{L}\mathcal{M}\mathcal{J}(\bar{\nu}_s, s) = \mathcal{L}\mathcal{I}\mathcal{J}(\bar{\nu}_s, \widetilde{\mathcal{M}}(\bar{\nu}_s)s) \tag{3.8}$$

where the linearized data is given by

$$\mathcal{L}d = \mathcal{I}\varphi_{,r}(\bar{\nu}_s, 0)\bar{r} \simeq d - \mathcal{I}\varphi(\bar{\nu}_s, 0) \simeq d. \tag{3.9}$$

Hence

$$\inf_{\tilde{r}} \mathcal{L}\mathcal{I}\mathcal{J}(\bar{\nu}_s, \tilde{r}) = \mathcal{L}\mathcal{I}\mathcal{J}(\bar{\nu}_s, \bar{\tilde{r}}) = 0 \tag{3.10}$$

with $\bar{\tilde{r}} = H^{-\frac{1}{2}}\bar{r}$.

But the linearized forward map $\mathcal{I}\varphi_{,\tilde{r}}(\bar{\nu}_s, 0)$ is not injective, hence there exists, besides $\bar{\tilde{r}}$, many $\tilde{r}$ such that $\mathcal{I}\varphi_{,\tilde{r}}(\bar{\nu}_s, 0)\tilde{r} = \mathcal{L}d$. Let us denote by $\bar{\tilde{r}}_{\min}$ the one of minimum norm. Necessarily, $\bar{\tilde{r}}_{\min}$ is in the orthogonal of the kernel of $\mathcal{I}\varphi_{,\tilde{r}}(\bar{\nu}_s, 0)$, *i.e.*

$$\bar{\tilde{r}}_{\min} \in \overline{\mathrm{Im}\mathcal{I}\varphi_{,\tilde{r}}(\bar{\nu}_s, 0)^T} \tag{3.11}$$

where the upper bar denotes the closure. Hence, for any $\varepsilon > 0$, there exists an $\bar{s}_\varepsilon$ such that

$$\|\bar{\tilde{r}}_{\min} - \mathcal{I}\varphi_{,\tilde{r}}(\bar{\nu}_s, 0)^T \bar{s}_\varepsilon\| \leq \varepsilon \tag{3.12}$$

i.e. , using (3.6),

$$\|\bar{\tilde{r}}_{\min} - \widetilde{\mathcal{M}}(\bar{\nu}_s)\bar{s}_\varepsilon\| \leq \varepsilon \tag{3.13}$$

which, because of (3.8), implies that

$$\mathcal{L}\mathcal{M}\mathcal{J}(\bar{\nu}_s, \bar{s}_\varepsilon) \leq C\varepsilon \tag{3.14}$$

where C is the norm of the forward linearized model map $\mathcal{I}\varphi_{,\tilde{r}}(\bar{\nu}_s, 0)$. This proves that

$$0 = \inf_{\tilde{r}} \mathcal{L}\mathcal{I}\mathcal{J}(\bar{\nu}_s, \tilde{r}) = \inf_{s} \mathcal{L}\mathcal{M}\mathcal{J}(\bar{\nu}_s, s). \tag{3.15}$$

3.1.3 FULL FORWARD MODEL: NUMERICAL RESULT

When the full non linear forward model $\mathcal{I}\varphi(\bar{\nu}_s, r)$ is used, we give now numerical evidence of the existence of $\bar{s}_\epsilon$ by minimizing $\mathcal{MJ}(\bar{\nu}_s, s)$ with respect to s. The Figure 3.1 shows the result of this minimization for the Synclay data set, described in Clément and Chavent (1992). Starting from the initial guess $s^0 = d - \mathcal{I}\varphi(\bar{\nu}_s, 0)$, the error function $\mathcal{MJ}$ decreases from 59409 to 29698 in 24 iterations (3 hours on the Cray-2). The resimulated data with $s^{24}(\bar{\nu}_s)$ is undistinguishable by the eye from the Synclay data d, so we shall call

$$\bar{s} = s^{24}(\bar{\nu}_s)$$

and now we can consider that

$$d = \varphi(\bar{\nu}, \bar{\sigma}) = \mathcal{I}\varphi(\bar{\nu}_s, \bar{r}) = \mathcal{M}\varphi(\bar{\nu}_s, \bar{s}).$$

To perform the optimization, we have used a Limited Memory Quasi-Newton Method for unbounded variables (Liu and Nocedal, 1989), described in the Appendix B.

In practice, of course, the true propagator $\bar{\nu}_s$ is unknown, and we illustrate on Figure 3.2 the important influence of the slowness when resimulating with the "true" time reflectivity $\bar{s} = s^{24}(\bar{\nu}_s)$. The four last seismograms represent the resimulations with $\bar{\nu}_s$, and uniform propagator models corresponding to 1500 m/s, 1800 m/s and 2500 m/s (in the Synclay data set, the velocity ranges from 1500 m/s to 4000 m/s, see Clément and Chavent, 1992). They have to be compared to the data (at the top). Of course, each event is correctly located in time: this is the role of the MBTT change of unknown. But the amplitudes obtained with the uniform propagators are really wrong, especially with 2500 m/s which is yet close to the mean value of the true propagator. The coupling between the modified unknowns (ν_s, s) is much less important than between the material unknowns (ν, σ) or between the intermediate unknowns (ν_s, r) (because the time shifts have desappeared), but the modified unknowns are not completely decoupled: minimizing $\mathcal{MJ}$ with respect to s with uniform propagator models would obviously not produce the same time reflectivity as $\bar{s}$. Hence, it is not worth minimizing with respect to the time reflectivity as long as the propagator is too wrong.

3.2 CASE OF MULTIPLE SHOTS

Now, we consider the legitimacy of the $s \mapsto r$ change of unknown in the case of multiple shots.

3.2.1 THE RESIDUALS: A GOOD FIRST GUESS OF $\bar{s}$

Of course, the collection

$$s^0 = (s_1^0, \ldots, s_{\text{NSHOT}}^0) \tag{3.16}$$

of the residuals associated to each shot

$$s_n^0 = d_n - \mathcal{I}\varphi_n(\bar{\nu}_s, 0) \simeq d_n. \qquad \text{for all } n = 1, \ldots, \text{NSHOT} \tag{3.17}$$

allows a reasonable shot-by-shot resimulation of the data $d = (d_1, \ldots, d_{\text{NSHOT}})$. But when the correct propagator $\bar{\nu}_s$ is used, the individual migrated sections $r_n = \mathcal{M}_n(\bar{\nu}_s)s_n^0$ will

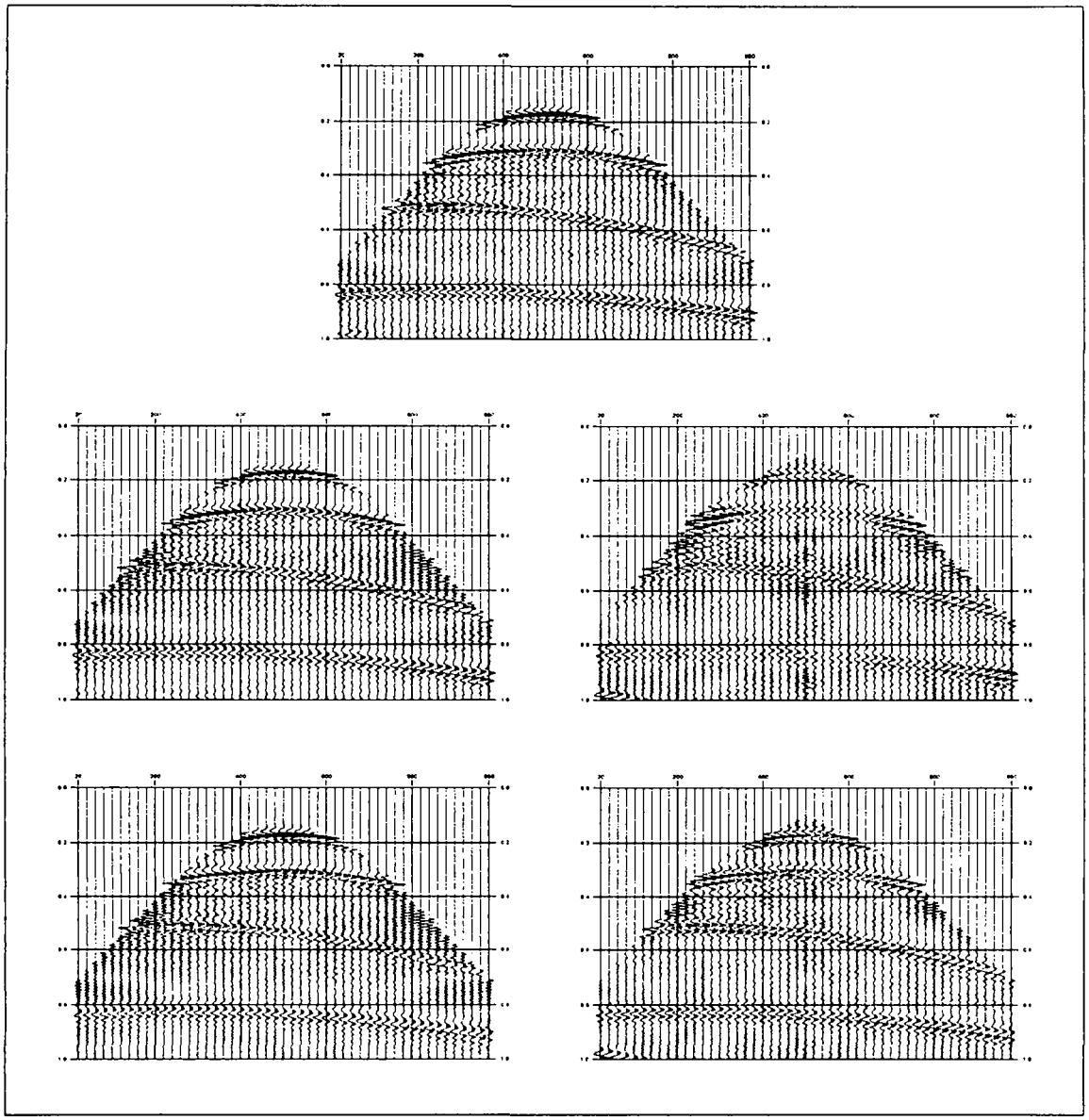

Fig. 3.1 Determination of the true time reflectivity $\bar{s}$: minimization with respect to s with the true propagator $\bar{\nu}_s$. Top: data $d = \varphi(\bar{\nu}, \bar{\sigma})$. Middle: initial time reflectivity $s^0 = d - \mathcal{I}\varphi(\bar{\nu}_s, 0)$ (residual), and the corresponding resimulation $\mathcal{M}\varphi(\bar{\nu}_s, s^0)$. Bottom: time reflectivity after 24 minimization iterations $s^{24}(\bar{\nu}_s)$, and the corresponding resimulation $\mathcal{M}\varphi(\bar{\nu}_s, s^{24}(\bar{\nu}_s))$.

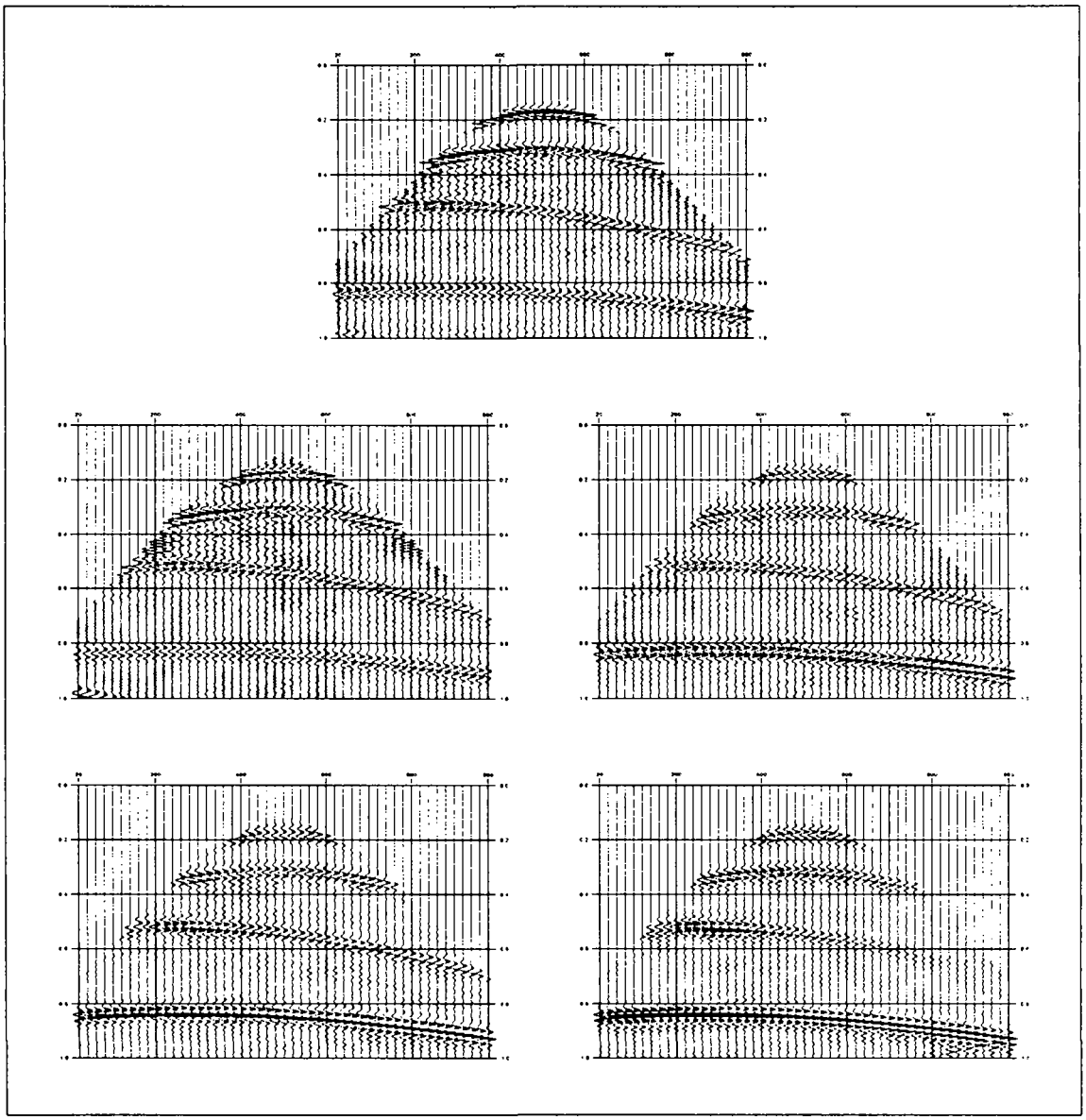

Fig. 3.2 Coupling between the propagator ν_s and the time reflectivity s. Top: data $d = \varphi(\bar{\nu}, \bar{\sigma})$. From left to right and from middle to bottom: resimulations with the true time reflectivity. First with the true propagator $\mathcal{M}\varphi(\bar{\nu}_s, \bar{s})$ then with uniform propagators $\mathcal{M}\varphi(1/1500, \bar{s})$, $\mathcal{M}\varphi(1/1800, \bar{s})$ and $\mathcal{M}\varphi(1/2500, \bar{s})$.

add constructively, so that the stacked section r will also allow a reasonable resimulation of the data of each shot. This shows that

> $s^0 \simeq d = (d_1, \ldots, d_{\mathrm{NSHOT}})$ is a reasonable approximation of the true time reflectivity $\bar{s}$ such that $(\bar{\nu}_s, \bar{s})$ and $(\bar{\nu}_s, \bar{r})$ are equivalent.

This is a great advantage of the $r \rightarrow s$ parametrization: we replace the search for $(\bar{\nu}_s, \bar{r})$, where both of them are unknown, by the search for $(\bar{\nu}_s, \bar{s})$, where $\bar{s}$ is already almost known from the data themselves.

As a practical consequence, we expect that a first determination of the background slowness ν_s will be possible by minimization of $\mathcal{MJ}(\nu_s, s^0)$ with respect to ν_s. A preliminary numerical result in this direction is given later below in this paper (*cf.* Figure 5.5).

3.2.2 LINEARIZED FORWARD MODEL: THEORETICAL RESULT

Suppose now that we use a linearized forward model

$$\mathcal{I}\varphi_{,r}(\nu_s, 0) = (\mathcal{I}\varphi_{1,r}(\nu_s, 0), \ldots, \mathcal{I}\varphi_{\mathrm{NSHOT},r}(\nu_s, 0)) \tag{3.18}$$

and that we replace the definition (2.4), (2.5) of the stacked migrated reflectivity r associated to the time reflectivity $s = (s_1, \ldots, s_{\mathrm{NSHOT}})$ by

$$r = \mathcal{M}(\nu_s)s \tag{3.19}$$

where $\mathcal{M}(\nu_s)$ is a stacked migration operator defined (compare to the definition (3.3) used for one individual shot) by

$$\mathcal{M}(\nu_s) = H\mathcal{I}\varphi_{,r}(\nu_s, 0)^T \tag{3.20}$$

where H is a symmetric positive definite operator in charge of restoring the "true amplitudes" of the stacked migrated section.

Then from (3.3), (3.18), (3.19) and (3.20) we see that the stacked migrated section r is now given by

$$r = H \sum_{n=1}^{\mathrm{NSHOT}} H_n^{-1} r_n \tag{3.21}$$

with

$$r_n = \mathcal{M}_n(\nu_s)s_n, \qquad \text{for all } n = 1, \ldots, \mathrm{NSHOT}. \tag{3.22}$$

Comparing (3.21), (3.22) and (2.4), (2.5) shows that these two definitions of the stacked reflectivity will coïncide when the same "true amplitude" operators H_n and H are used for each individual shot and for the stacked section. This is often the case, as usually H_n and H are chosen, in a first approximation, to depend only on the depth z. But in the general case, the definition (3.19), (3.20) or, equivalently, (3.21), (3.22) has to be prefered to (2.4), (2.5). Notice also that (3.21), (3.22) is easier to implement than (2.4), (2.5), as it rewrites

$$r = H \sum_{n=1}^{\mathrm{NSHOT}} \mathcal{I}\varphi_{n,r}(\nu_s, 0)^T s_n \tag{3.23}$$

which shows that it is not necessary to know the individual "true amplitude" factors H_n of each shot, but only the factor H for the stacked section.

We can now proceed exactly as in the single shot case: consider the linearization at $\nu_s = \bar{\nu}_s$ and all formulas (3.5) through (3.15) can be re-read in the multiple shot context, by replacing references to (3.3) by references to (3.20). This proves the existence, for any $\varepsilon > 0$, of an $\bar{s}_\varepsilon = (\bar{s}_{\varepsilon 1}, \ldots, \bar{s}_{\varepsilon \mathrm{NSHOT}})$ such that

$$\mathcal{LMJ}(\bar{\nu}_s, \bar{s}_\varepsilon) \leq C\varepsilon \tag{3.24}$$

so that

$$0 = \inf_{\tilde{r}} \mathcal{LIJ}(\bar{\nu}_s, \tilde{r}) = \inf_{s=(s_1,\ldots,s_{\mathrm{NSHOT}})} \mathcal{LMJ}(\bar{\nu}_s, s). \tag{3.25}$$

Hence the $r \rightarrow s$ change of unknowns is fully legitimated in the linearized multiple shot case, provided that stacking formula (3.21), (3.22) or, equivalently, (3.23) is used.

4 MINIMIZATION WITH RESPECT TO THE SLOWNESS BACKGROUND IN THE SINGLE SHOT CASE

The usual objective function $\mathcal{J}(\nu, \sigma)$ is known to possess, already in the case of a single shot, many local minima with respect to ν (for fixed σ), which prevent local optimization methods to find the global minimum: they usually stop in a local minimum, where the corresponding synthetic seismogram is still far from the data.

We perform now the optimization of the modified function $\mathcal{MJ}(\nu_s, s)$ with respect to ν_s, to show that convergence to a global minimum can be achieved, meaning that the MBTT change of unknown has a "smoothing" effect on the objective function, and thus convergence of a local optimization method will not be attraped in a bad local solution. We have already presented (Clément and Chavent, 1992) preliminary evidence of this effect using a 2D cross-section in the ν_s-space.

To optimize with respect to ν_s, we have also used a Limited Memory Quasi-Newton Method for bounded variables (Byrd *et al.*, 1993) and a brief description of the method is also given in Appendix B. The problem is now bounded because the velocity has to be strictly positive and less than a maximum value (in order to satisfy the stability condition of the numerical scheme for the wave equation).

4.1 MULTISCALE MINIMIZATION

We present in Figures 4.1, 4.2 and 4.3 the results of the multiscale minimization with respect to the propagator unknown with the true time reflectivity $\bar{s}$. We represent the velocity instead of the slowness because its values are more meaningful.

In the three figures, the top raw shows the target propagator and the data for the Synclay model (see Clément and Chavent, 1992). In the middle of Figure 4.1, there is the uniform initial guess of 2500 m/s for the velocity and the corresponding seismogram—which is far from the data ! Then we have let the minimization method run for 3 hours (Cray-2) for each scale. That is, we first minimize on the coarsest grid (0th scale: 4×3 grid = 12 degrees of freedom) using the initial point mentioned before and obtain, after 3 hours on a Cray-2, $\nu_{s0}^\star$. And then, taking this value as the initial point, we minimize on the next finer grid (1st scale: 7×5 grid = 35 d. of f.) and obtain $\nu_{s1}^\star$, and go on in this fashion to minimize on the following grids (2nd scale: 13×9 grid = 117 d. of f., 3rd

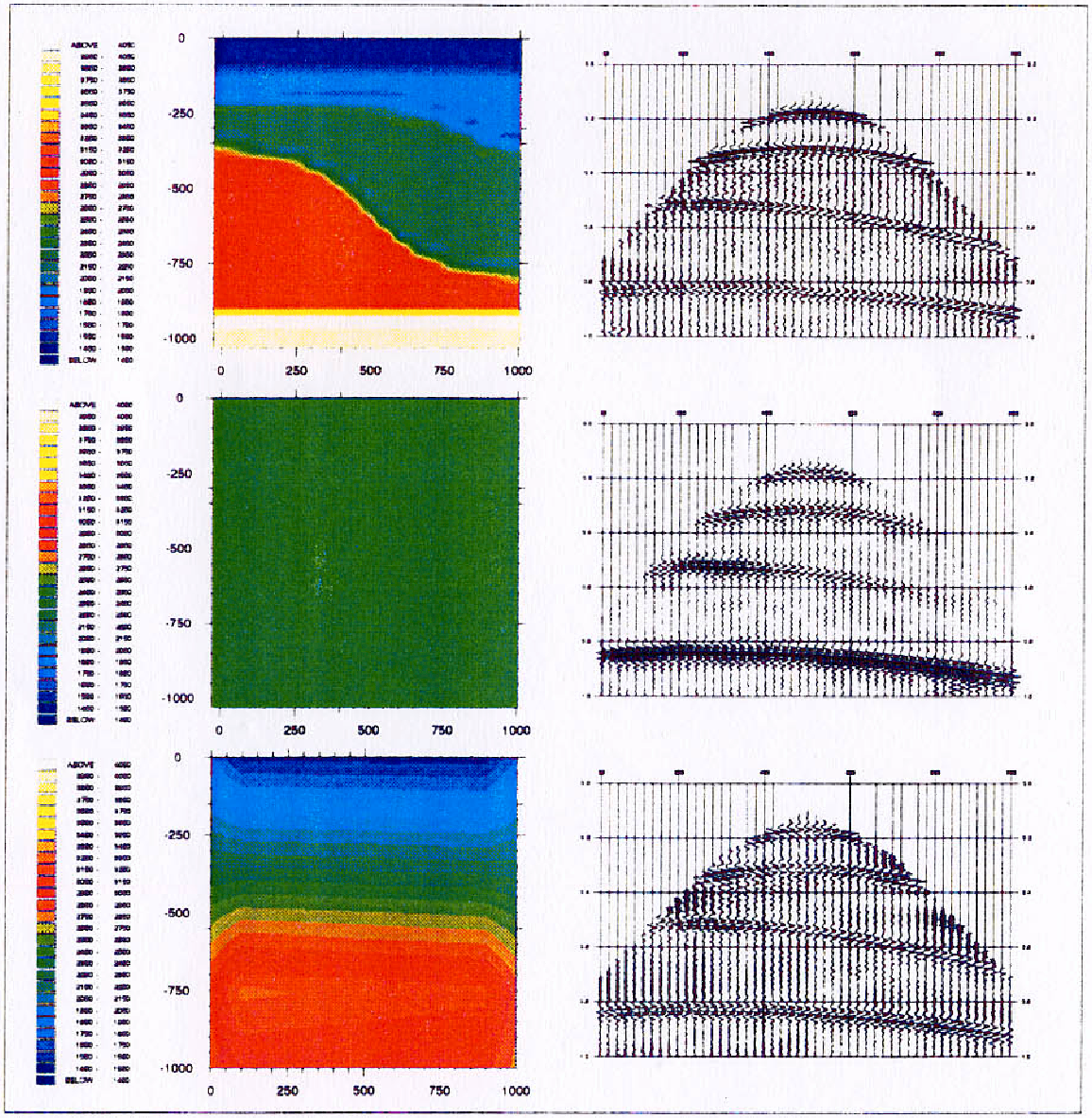

Fig. 4.1 Multiscale minimization with respect to v, with the true time reflectivity $s(t)$. Top: true velocity $\bar{v}$, and data $d = \varphi(v, \sigma) = \mathcal{M}\varphi(\bar{v}, s)$. Then, left : velocities, and right: corresponding resimulations. Middle: uniform initial guess 2500 m/s. Bottom: after 3 hours of minimization on a Cray-2 on the 0th scale from the initial guess.

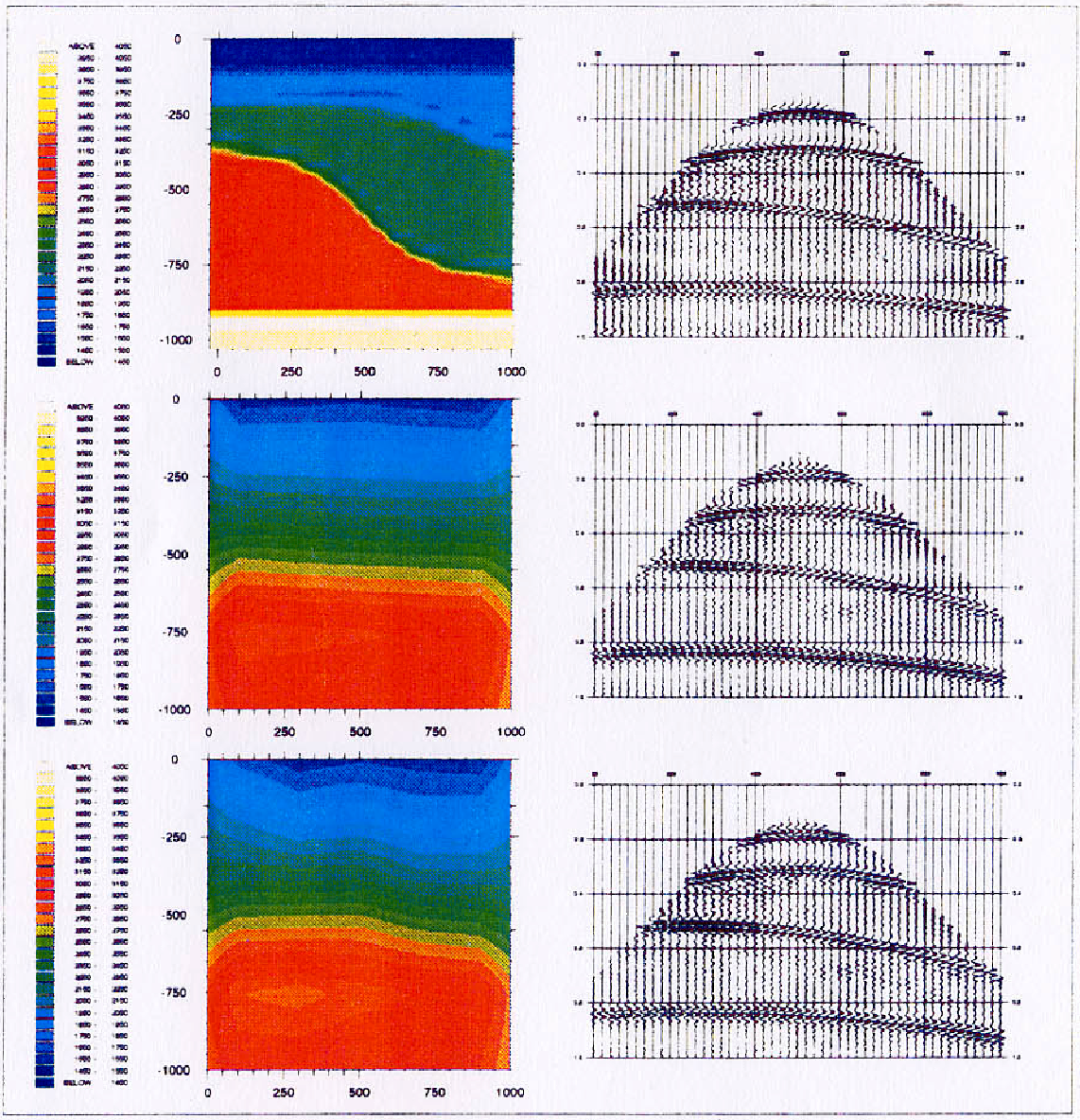

Fig. 4.2 Multiscale minimization with respect to v, with the true time reflectivity $s(t)$. Top: true velocity v_{true} and data $d = \varphi(v, \sigma) = \mathcal{M}\varphi(v_{\text{true}}, s)$. Then, left : velocities, and right: corresponding resimulations. Middle: after 3 hours of minimization on a Cray-2 on the 1st scale from the minimum on the 0th scale. Bottom: after 3 hours of minimization on a Cray-2 on the 2nd scale from the minimum on the 1st scale.

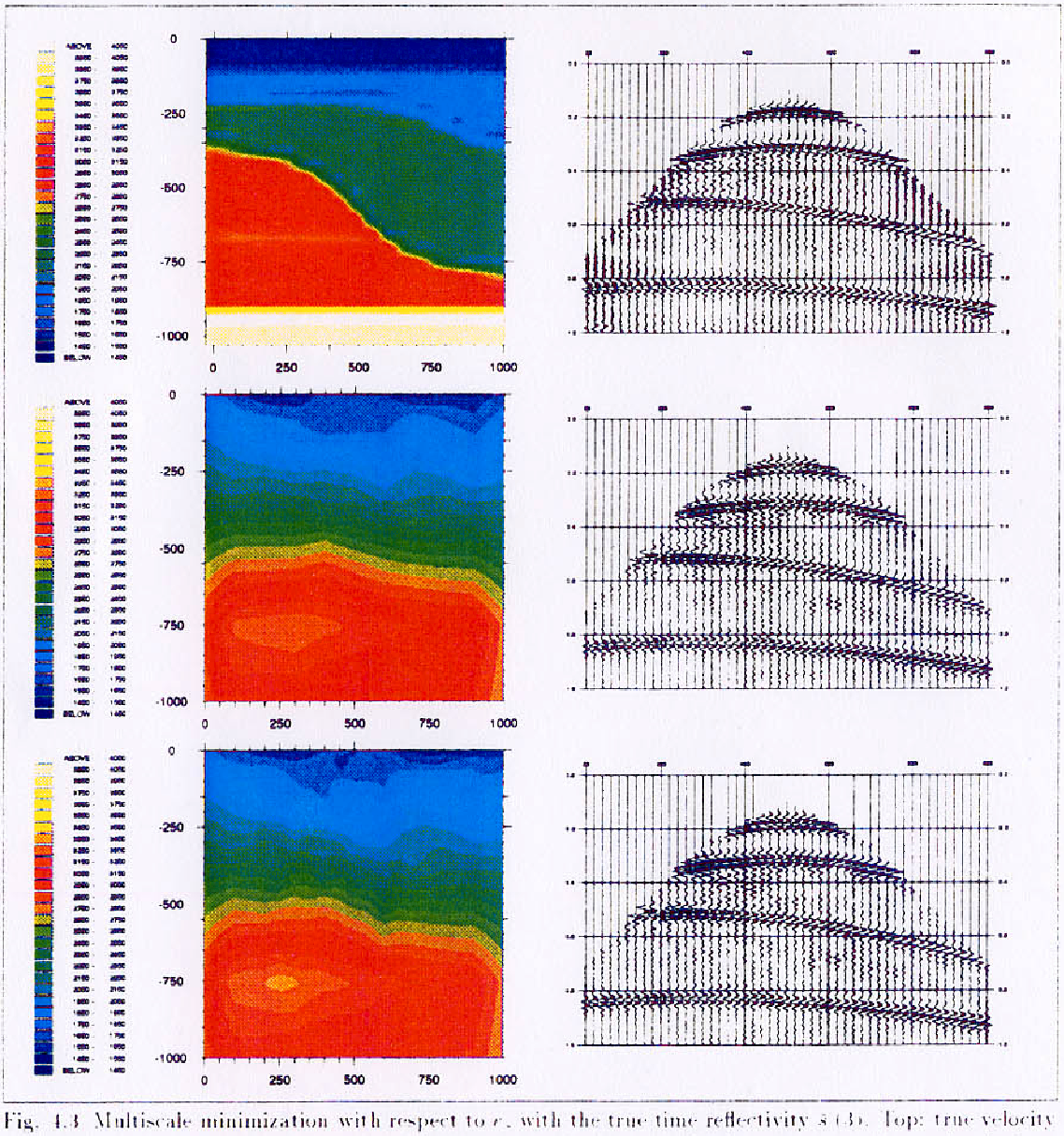

Fig. 4.3 Multiscale minimization with respect to c, with the true time reflectivity s (3). Top: true velocity c_{true} and data $d = \varphi(c, \sigma) = \mathcal{M}\varphi(c_{\text{true}}, s)$. Then, left : velocities, and right: corresponding resimulations. Middle: after 3 hours of minimization on a Cray-2 on the 3rd scale from the minimum on the 2nd scale. Bottom: after 3 hours of minimization on a Cray-2 on the 4th scale from the minimum on the 3rd scale.

scale: 25×17 grid = 425 d. of f. and 4th scale: 49×33 grid = 1617 d. of f.) using always as the initial point the approximated minimum found on the previous scale. The total minimization time was of 15 hours of Cray-2. The corresponding seismograms are getting really close to the data and we can conclude that we have reached a global minimum.

The results are surprisingly good, considering that we have only used here single shot data: already after the minimization on the 0th scale, a correct information on the general trend of the true velocity has raised. This is certainly due to the fact that although we have started from a very poor initial guess—a uniform model—its value was closed to the average of the true velocity model.

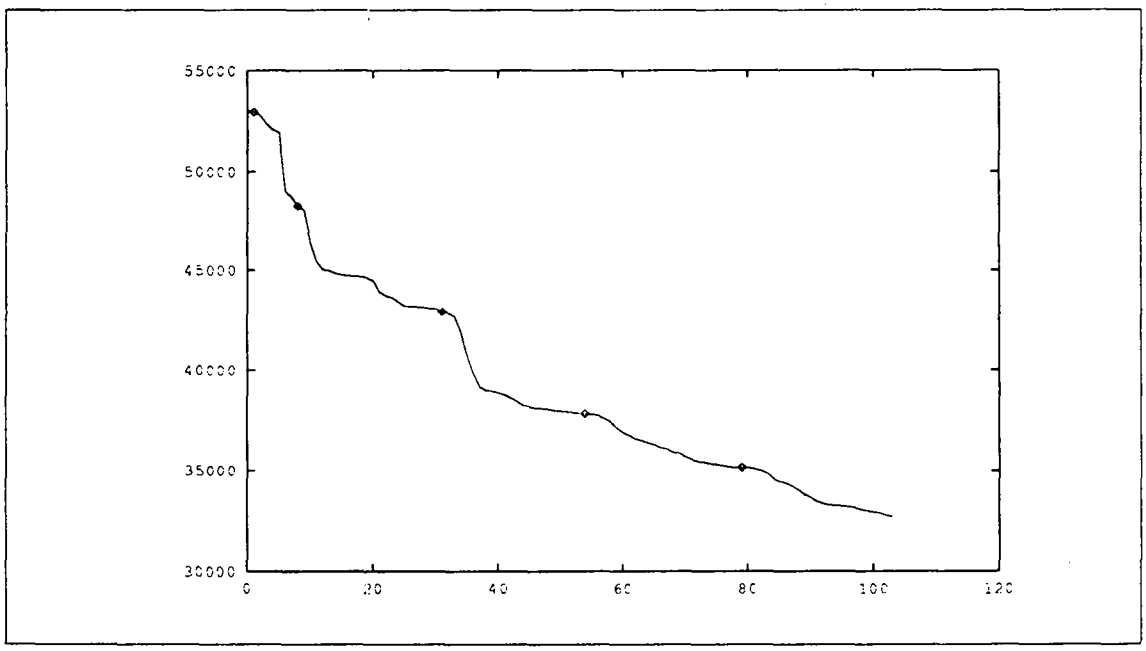

Fig. 4.4 Decrease of the modified data misfit function during the multiscale minimization. Horizontal axis: number of minimization iterations. Vertical axis: value of the modified data misfit function $\mathcal{MJ}$. *Diamonds: first iteration of each scale level (from 0 to 4).*

In Figure 4.4, we show the decrease of the modified criterion $\mathcal{MJ}$ during its hierarchical minimization with respect to the propagator. The beginning for each scale, from level 0 to level 4, is symbolized by a diamond. Hence each diamond indicates the beginning of 3 hours of minimization. It is then important to notice that the main gain is realized during the minimization on the coarsest grid which is also the most expensive: only 8 iterations during the 3 hours against 23 or 25 iterations for all the other scales during the same 3 hours. But the most interesting result concerns the hierarchical organization of the propagator: at least for the three first scale levels, an important decrease occurs just after considering this scale level. This means that it is efficient to increase the number of degrees of freedom progressively. And it gives an *a posteriori* justification of the moderate overcost brought by the multiscale representation of the propagator.

4.2 NON UNIQUENESS OF THE GLOBAL MINIMUM

Of course, because we consider here only single shot data, there is not a unique global

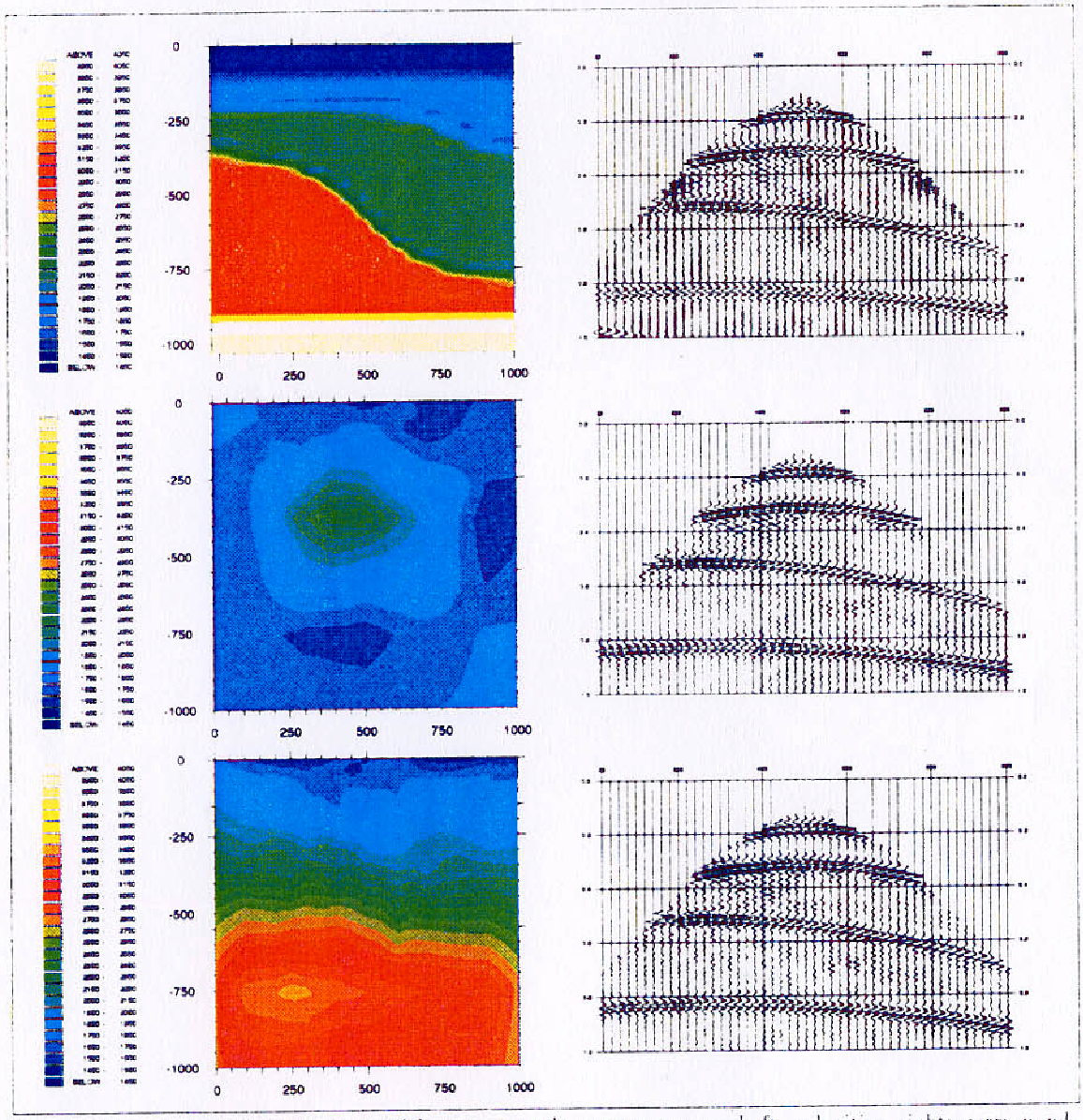

Fig. 4.5 No unique global minimum with respect to the propagator e_-. Left: velocities, right: corresponding resimulations with the true time reflectivity s. Top: true velocity. Middle: after 15 hours of multiscale optimization on a Cray-2 from initial guess 1800 m/s. Bottom: after 15 hours of multiscale optimization on a Cray-2 from initial guess 2500 m/s.

minimum. This is illustrated in Figure 4.5 where we show three different velocity models which all produce a seismogram that fits well the data (when resimulated with the true time reflectivity $\bar{s}$). They are the true propagator (top) and two results of multiscale minimization (3 hours per scale) from uniform initial guesses of 1800 m/s (middle) and 2500 m/s (bottom). The last one being the result of the experiment described before. The velocity in the middle is certainly not a correct approximation of the true one—although it reproduces well the data—and we can hope that such global solutions will be dropped out when we will handle multishot data.

All the minimization experiments we have run never converged to bad local solutions so we state that the smoothing effect brought by the MBTT change of unknown—elimination of the phase shifts—apparently worked well and our modified data misfit function is minimizable by (yet sophisticated) local methods.

5 EXTRACTING INFORMATION ON THE BACKGROUND SLOWNESS IN THE MULTIPLE SHOTS CASE

In the previous section, we showed that our modified criterion $\mathcal{MJ}$ was suited for the minimization with respect to the propagator unknown by a local method; but when limited to the single shot case, it subsisted, of course, the problem of the under-determination of the unknowns and it existed several global minima for the propagator.

Multiplying the shots is the common solution to remove this under-determination and the difficulty is then to build a single Earth model from this multiplicity of measures. We have focused here on the simplest "stacking" version of our algorithm (see section 2).

5.1 MULTISHOT MBTT ALGORITHM

We illustrate now the effect of a change of propagator on the multishot MBTT algorithm, and, in particular, how the intermediate depth reflectivity is distorded when a wrong propagator is used.

The three Figures 5.2, 5.3 and 5.4 follow exactly the same pattern: on the top, the chosen propagator, then, in the middle, the intermediate stacked depth reflectivity computed from the chosen propagator and the corresponding residuals as the time reflectivities (on the left, the stacked depth reflectivity with respect to the velocity and, on the right, the stacked depth reflectivity with respect to the impedance). Finally, on the bottom, the resimulation obtained after having updated the smooth backgrounds ν_s and σ_s with these depth reflectivities. This resimulation has to be compared with the 5-shots Synclay data in Figure 5.1. The difference between the three resimulations is the initial propagator which is the true propagator in Figure 5.2 and uniform propagators in Figure 5.3 (1500 m/s) and in Figure 5.4 (1800 m/s).

Of course, when the true propagator is used, each individual migrated section $r_n = \mathcal{M}_n(\bar{\nu}_s)s_n^0$ gives a correct partial image of the Earth. Thus, they all add constructively into the stacked section r which allows a reasonable resimulation of the data for each shot. We just notice some loss of signal in the extreme parts for the extremes shots: where the Earth is less illuminated.

On the contrary, when a very poor propagator guess is used (1500 m/s), each individual

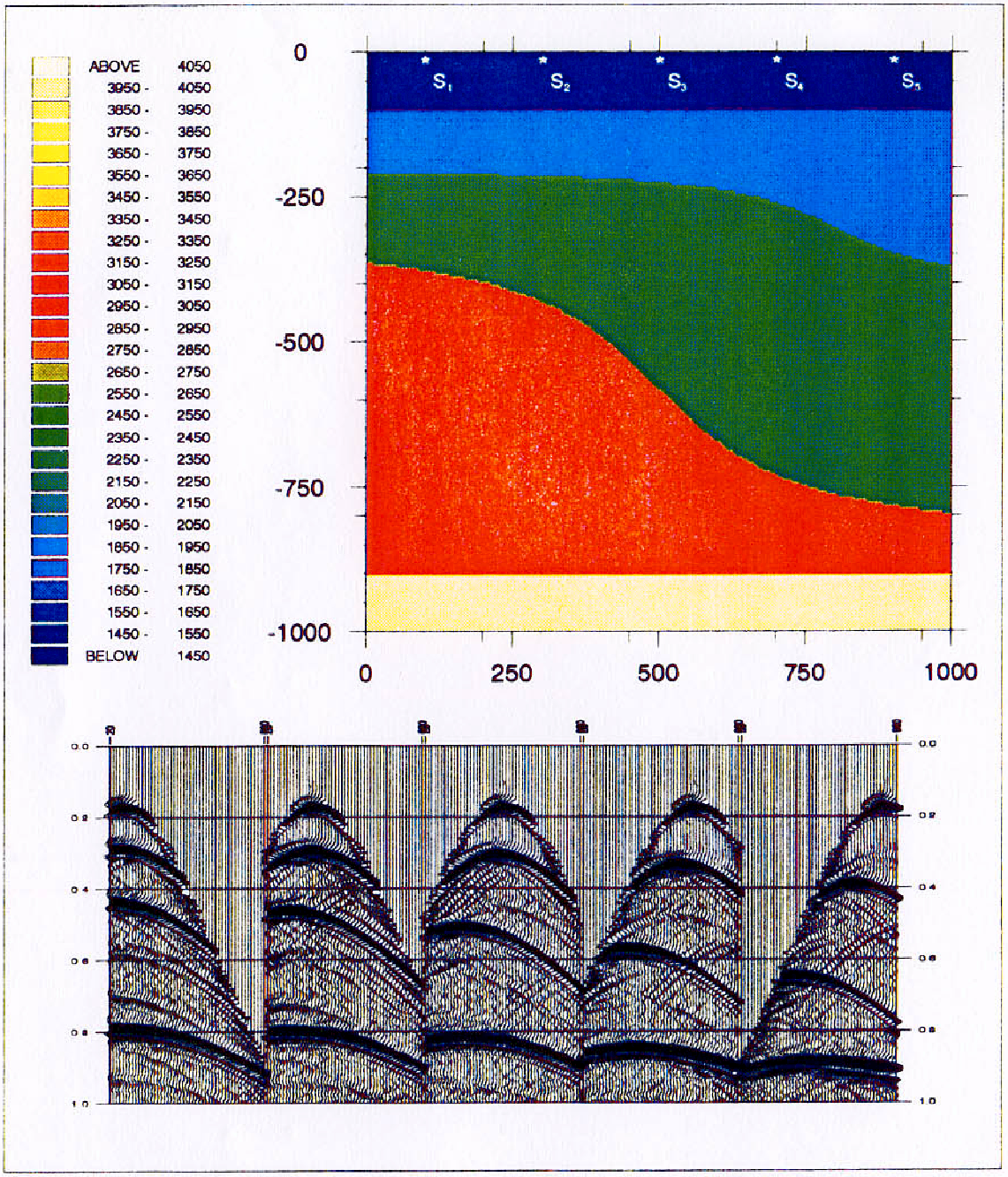

Fig. 5.1 Multishot data. Top: Synclay velocity model with five layers (1500 m/s, 1800 m/s, 2300 m/s, 3000 m/s and 4000 m/s). Uniform density $\rho = 1000$ kg/m^3. Bottom: seismograms corresponding to the five shots located at 100 m, 300 m, 500 m, 700 m, 900 m.

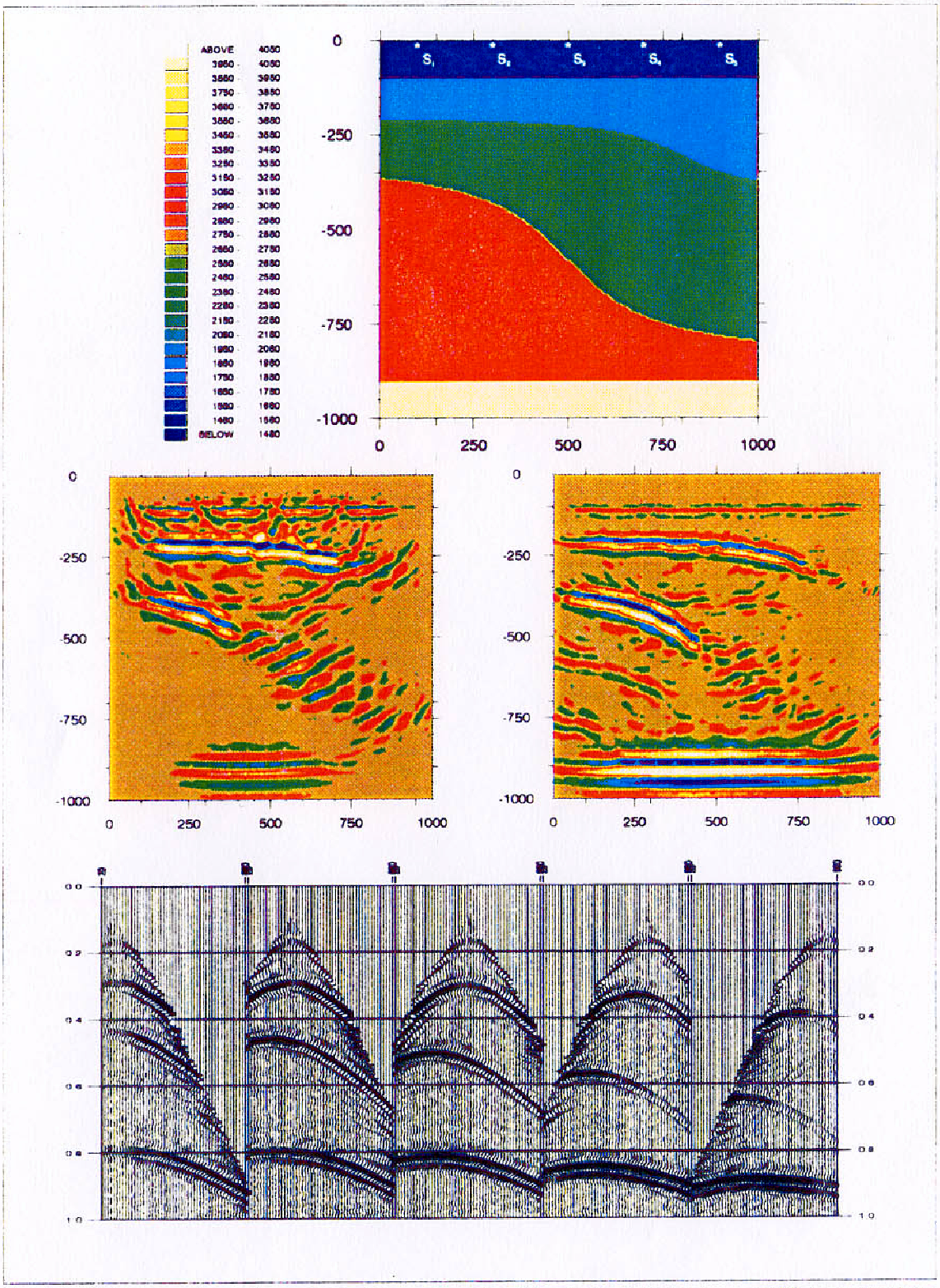

Fig. 5.2 Multishot resimulation with the true propagator. Top: true propagator c. Middle: intermediate stacked depth reflectivity with respect to c (left) and σ (right). Bottom: corresponding resimulation.

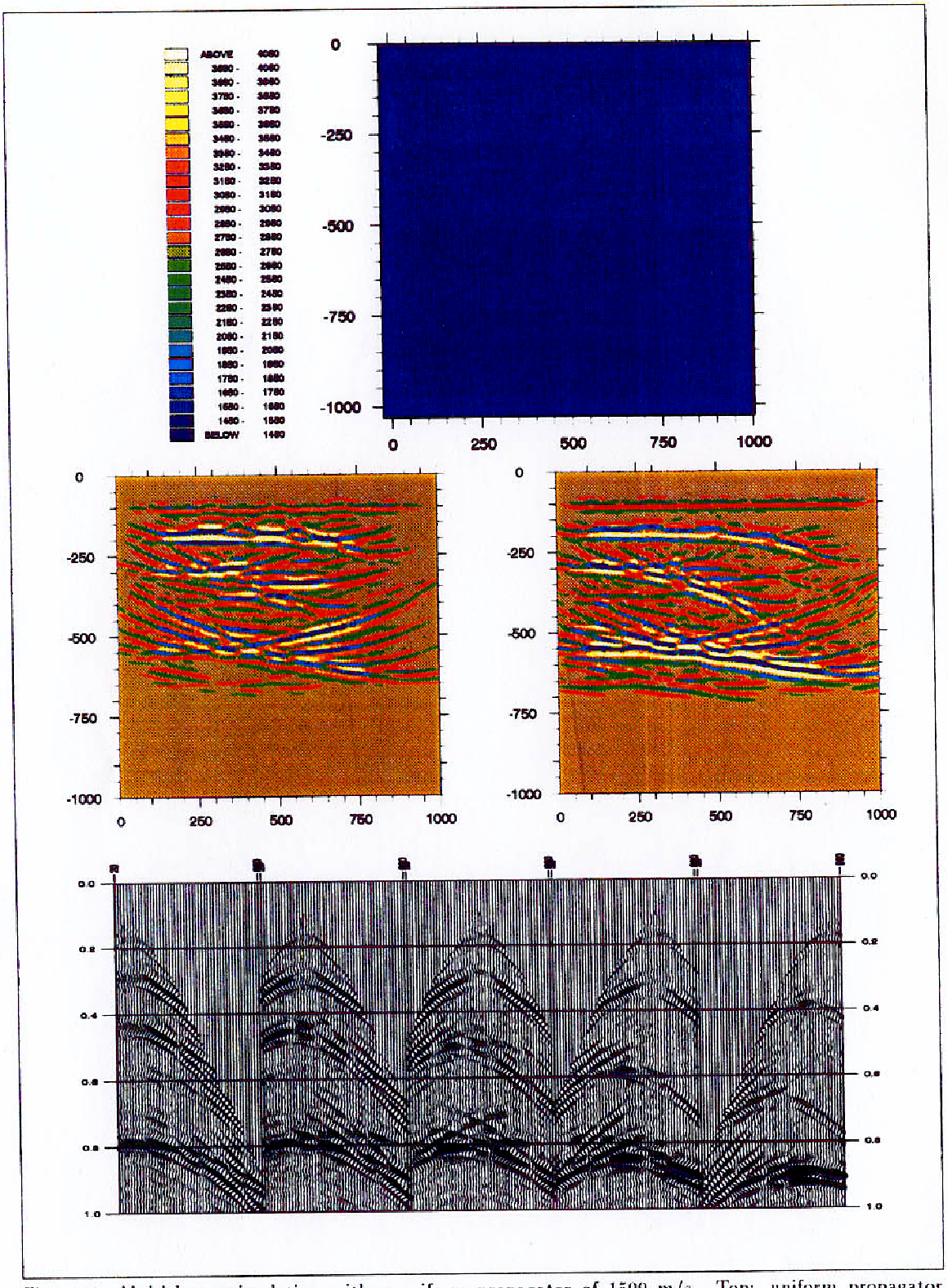

Fig. 5.3 Multishot resimulation with a uniform propagator of 1500 m/s. Top: uniform propagator 1500 m/s. Middle: intermediate stacked depth reflectivity with respect to c (left) and σ (right). Bottom: corresponding resimulation.

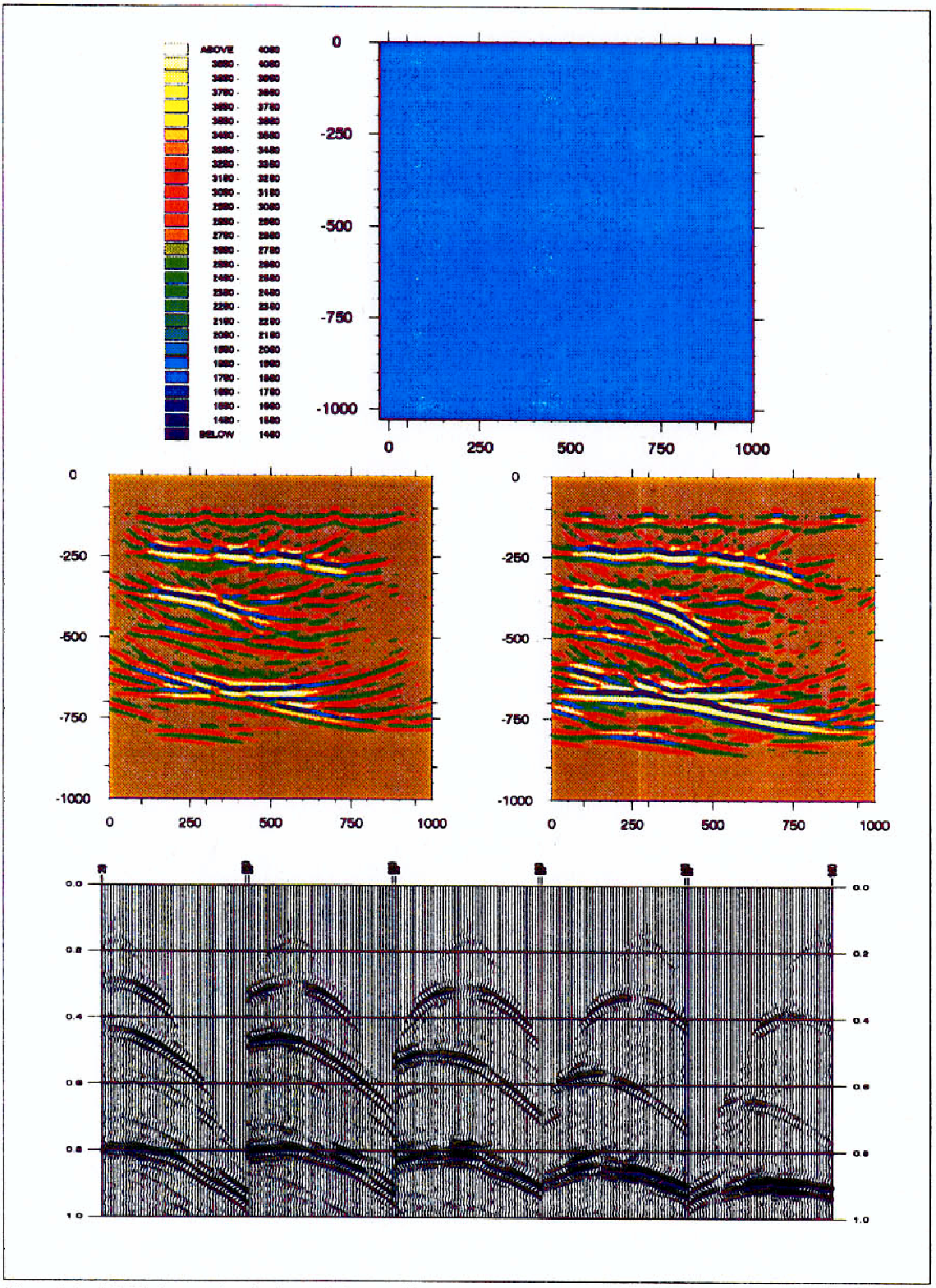

Fig. 5.4 Multishot resimulation with a uniform propagator of 1800 m/s. Top: uniform propagator 1800 m/s. Middle: intermediate stacked depth reflectivity with respect to c (left) and σ (right). Bottom: corresponding resimulation.

migrated section gives a different partial image of the Earth. So, the stacked section is a really fuzzy image. And the resimulation show losses of signal even in fully illuminated areas. This comment is also valid for the case of 1800 m/s in spite of the feeling of clearness which comes out of the corresponding resimulation.

This establishes the sensibility of the MBTT modified criterion to the propagator unknown in the case of 5-shots data. The next step is to examine the shape of its gradient with respect to the propagator.

5.2 GRADIENT

Since we have only considered, up to now, a simple stacking version of the MBTT algorithm for the multishot case, computing the gradient of the modified data misfit function $\mathcal{MJ}$ with respect to its propagator argument ν_s amounts only to stack the contributions of all single shots given in Clément and Chavent (1992).

The computation of this gradient has successfully passed the Finite Differences Test (see Appendix A). It was a *sine qua non* condition for further minimization experiments.

In Figure 5.5, we present our first gradient calculation in the multishot case. The chosen propagator guess is a uniform model of 1800 m/s (top right). When compared to the true propagator model (top left), it is too high in the first hundred meters and too low under 250 m. The next five charts show the gradient of $\mathcal{MJ}$ with respect to the propagator projected on the grids of different level: from the 0th scale (grid cell of 800 m) to the 4th scale (grid cell of 50 m). This is not a minimization experiment but just the first gradient for which we show gradually the information it contains. And this first step is very promising.

It has to be compared to the velocity model of Figure 4.5 in the middle left, which corresponds to more than a hundred minimization iterations from the same uniform initial guess of 1800 m/s but with single shot data. In that case, the minimization method was not able to retrieve the general shape of the true propagator and stayed in another global minimum (the shot gather was correctly fitted). Here, already at the first iteration, the gradient is pointing into the correct direction. The warm colours represent positive values (where the propagator will decrease) and the cold colours are for the negative values (where the propagator will increase), the brown being for almost zero values. These colours are correctly located and, furthermore, the isolines of the gradient try to fit those of the true propagator (projection on the 2nd scale, lower middle right).

We can conclude that, even in its simple stacking version, the multishot MBTT algorithm shows its ability to extract information on the background slowness and this already after the first gradient step. This is very encouraging for further minimization experiments.

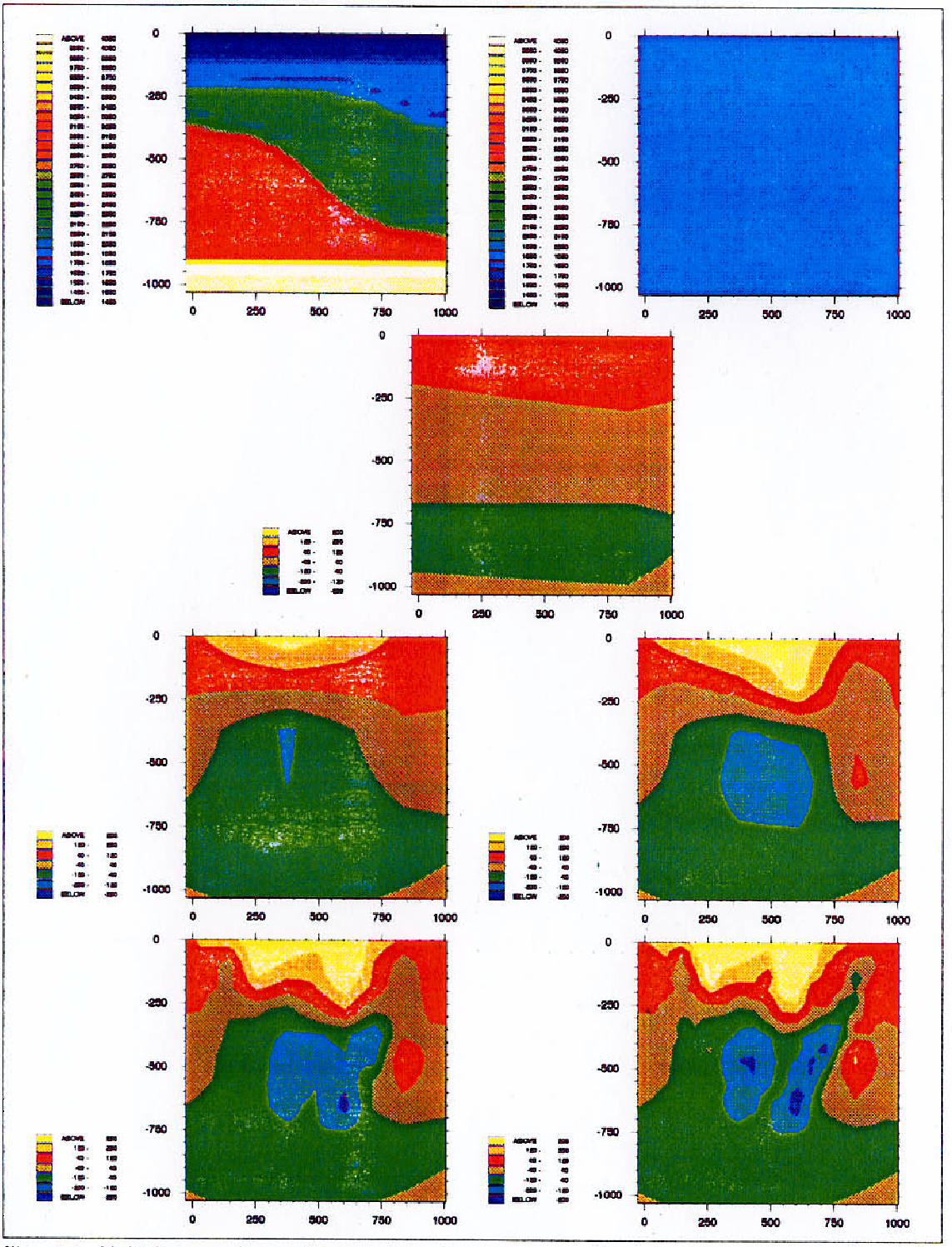

Fig. 5.5 Multishot: gradient with respect to the propagator. Top left: true propagator. Top right: uniform initial propagator 1800 m/s. Then, from left to right and from upper middle to bottom: gradient of the modified data misfit function $\mathcal{MJ}$ with respect to the propagator successively projected on the 0th, 1st, 2nd, 3rd and 4th scale grids.

6 CONCLUSION

The Migration-Based Travel Time (MBTT) approach to full waveform acoustic inversion eliminates the phase shifts on the synthetics which are responsible for the presence of local minima for the usual Output Least Square (OLS) data misfit function. Relying on two successive changes of unknowns from the OLS formulation, it allows the computation of a gradient with a reasonable overcost. This opens the way to minimization by local methods.

Our major findings are:

- The MBTT change of reflectivity unknown is legitimate: we are able to exhibit a "true" time reflectivity $\bar{s}$ which reproduces well our noiseless data—*i.e.* such that $\mathcal{MJ}(\bar{\nu}_s, \bar{s}) \simeq 0$. This result is proved theoretically when a linearized forward model is used (existence of a minimizing sequence with zero minimum for $\mathcal{MJ}$), and only numerically for the full forward model in the single shot case (minimization by a Quasi-Newton local method).

- In the single shot case, our modified criterion $\mathcal{MJ}$ is amenable to local minimization with respect to the smooth slowness ν_s without encountering bad local minima. But, of course, in that case there subsists several global minima—*i.e.* that all fit well the data. Furthermore, the multiscale parametrization of our propagator unknown ν_s allows an efficient hierarchical minimization.

- In the multiple shots case, already the first gradient is pointing into the correct direction in the propagator space.

These results are very satisfying, and we are now proceeding to the implementation of the minimization in the case of multiple shots.

ACKNOWLEDGMENTS

This research was carried out as part of the Prestak Structural Interpretation consortium project (PSI). The authors hereby acknowledge the support provided by the sponsors of this project.

REFERENCES

Bunks, C., 1993, Paraxial wave equation inversion of seismic data by migration and remodeling: submitted to Geophysical Prospecting.

Byrd, R. H., Lu Chen, P., and Nocedal, J., 1993, Bound constrained nonlinear optimization and limited memory methods: Technical Report, Northwestern University, Evanston, IL, to appear.

Chavent, G., and Jacewitz, C. A., 1990, Automatic determination of background velocities by multiple migration fitting: Proc. of the 60th annual Int. SEG Meeting.

Chavent, G., 1993, A time domain derivation of the Kirchhoff migration as the gradient of a data misfit function: Technical Report 1928, INRIA, Paris, France.

Clément, F., 1991, A migration-based travel time formulation for the inversion of 2D velocity distributions: PSI 1992 Annual Report, IFP, Rueil Malmaison, France.

Clément, F., and Chavent, G., 1992, Waveform inversion through MBTT formulation: PSI 1992 Annual Report, IFP, Rueil Malmaison, France.

Clément, F., and Chavent, G., 1993, Waveform inversion through MBTT reformulation: Proc. of the 2nd Int. Conf. on Math. and Num. Aspects of Wave Propagation, SIAM, Philadelphia, PA.

Gilbert, J. C., and Lemaréchal, C., 1989, Some numerical experiments with variable storage Quasi-Newton algorithms: Math. Programming, **45**(3), 407–435.

Liu, D. C., and Nocedal, J., 1989, On the limited memory BFGS methods for large scale optimization: Math. Programming, **45**(3), 503–528.

Nocedal J., 1980, Updating Quasi-Newton matrices with limited storage: Math. of Comp., **35**(151), 773–782.

Symes, W. W., and Kern, M., 1992, Velocity discrimination by differential semblance: Proc. of the 54th EAEG Meeting.

APPENDIX A: THE GRADIENTS OF $\mathcal{MJ}$

As we intend to use local optimization methods, and more precisely Quasi-Newton type algorithms, it is of prime necessity to have a correct computation of the gradient of the function that we want to minimize.

The function $\mathcal{MJ}$ is already complicated, and thus its gradient is much more complicated, but our code has finally passed a significant test: the finite differences check.

A.1 COMPUTATION

	wave eq.	cpu time	
$\mathcal{J}$	1	1	
$\nabla\mathcal{J}$	2	3.5	
$\mathcal{MJ}$	3	4.5	1
$\nabla\mathcal{MJ}$	6	16	3.5

Tab. A.1 Relative complexities of the usual and modified error functions and their gradients in terms of wave-equations-to-solve and of cpu time (Cray-2).

The MBTT algorithm, described in Clément and Chavent (1992), shows the complexity of the function $\mathcal{MJ}$ itself: the first step (migration) involves the resolution of two wave equations plus two correlations of the wavefields, the second step (update) involves technical steps as projection, weighting, normalization and then, the last two steps correspond

to the usual function $\mathcal{J}$. In terms of cpu time, the modified data misfit function $\mathcal{MJ}$ is 4.5 times more expensive than the usual one (see Table A.1). And its gradient $\nabla\mathcal{MJ}$ is again 3.5 times more expensive ! This is because of the complexity of the right-hand sides of the three adjoint wave equations and of the additional correlations, see Clément and Chavent (1992) for a completely detailed calculation of the gradients.

Thus, the complexity of the MBTT concept which appears in the computation of the function $\mathcal{MJ}$ is amplified when considering its gradient. The sources for errors in the implementation are numerous and we need tests to be sure of the correctness of the code.

A.2 VALIDATION OF THE CODE

First, one can check that the computed gradient defines, at least, a descent direction for the function. But the definitive test must be the *finite differences (FD) check:* at any grid point, the calculated gradient must agree with the FD approximations of the partial derivative of the function at that point. This test insures that one has computed the gradient corresponding to the computed function (it does not prove, of course, that the function is correctly computed !).

Let $\mathbf{x}$ be the unknown, distributed on a grid G,

$$\mathbf{x} = (\mathbf{x}_M)_{M\in G},$$

when $\mathbf{x}$ is the time reflectivity s, G corresponds to the time domain $\mathcal{G}\times[0,T]$ and when $\mathbf{x}$ is the slowness ν_s, G is the grid corresponding to the "smooth scales" (Clément and Chavent, 1992). For a given value $\mathbf{x}_0$ of the unknown, let $\nabla = (\nabla_M)_{M\in G} = \left(\frac{\partial \mathcal{MJ}}{\partial \mathbf{x}_M}(\mathbf{x}_0)\right)_{M\in G}$ be the gradient at that value and let us denote by ∇^{AS} and ∇^{FD} the gradients computed through the adjoint state method and through finite differences. For any point $P \in G$,

$$\nabla_P^{\mathrm{FD}} = \frac{\mathcal{MJ}(\mathbf{x}_0 + h^P\|\mathbf{x}_0\|) - \mathcal{MJ}(\mathbf{x}_0)}{h\|\mathbf{x}_0\|}$$

with $h^P = \left(h_M^P\right)_{M\in G}$ and $h_M^P = h\delta_{MP}$ $(h \in \mathbb{R})$.

Properties of the true gradient ∇ define possible tests for ∇^{AS}. The descent direction test can correspond to the fact that

$$\lim_{h\to 0} \frac{\mathcal{MJ}(\mathbf{x}_0 + h\nabla) - \mathcal{MJ}(\mathbf{x}_0)}{h} = \|\nabla\|^2$$

and the FD check to

$$\lim_{h\to 0} \nabla_P^{\mathrm{FD}} = \nabla_P, \quad \text{for any } P \in G.$$

Numerically, because of computer rounding errors, we cannot reproduce exactly the limit $h \to 0$ but we will be satisfied when the "stabilized decimals" of ∇_P^{FD} agree with those of ∇_P^{AS}. This is the case for all the tested points $P \in G$ (see an example in Table A.2) but only in the case of Dirichlet boundary conditions ! We postponed the search for the error in the boundary conditions of the adjoint equation, and this forced us to consider a grid which is four times larger and, of course, it penalizes the cpu time by the same factor of 4...

adjoint state	9.2120153
FD, $h = 1$	8.5322569
FD, $h = 10^{-1}$	9.1804073
FD, $h = 10^{-2}$	9.2034511
FD, $h = 10^{-3}$	9.2122295
FD, $h = 10^{-4}$	9.2120380
FD, $h = 10^{-5}$	9.2120437
FD, $h = 10^{-6}$	9.2123479
FD, $h = 10^{-7}$	9.2117115
FD, $h = 10^{-8}$	9.2573463

Tab. A.2 Decimals of the gradient of $\mathcal{MJ}$ with respect to the fourth scale of the velocity at the point ($x = 500$ m,$z = 750$ m) for the Synclay model. First line: the value calculated through the adjoint state method. Following lines: the values calculated by finite differences for decreasing steps h. The values are in agreement, at the precision of the calculation.

APPENDIX B: THE LIMITED MEMORY BFGS METHOD

In this work we want to solve minimization problems like

$$\inf_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x}) \tag{B.1}$$

where f is the least squares error function defined before in this paper, the variable $\mathbf{x} \in \mathbb{R}^n$ is either the propagator ν_s or the time reflectivity s and the domain $\mathcal{X}$ is either bounded or not. The dimensionality of the problem, in the single shot case, goes from 12 to 1617 bounded variables (for the propagator ν_s on a multiscale basis with 5 scale levels), and up to 61300 unbounded variables (for the time reflectivity s). In the multishot case, dimensionality of the time reflectivity s will increase tremendously and so we are dealing with a veritable large scale problem.

For large scale optimization problems with or without bounds, a classical Quasi-Newton method can be prohibited because of its large memory requirement, namely $O(n^2)$—*i.e.* 4 10^8 for s ! To solve these problems then, one can use a Limited Memory method based on the Quasi-Newton principle of using the change in the gradient to generate an approximation to the Hessian. For the limited memory method, this is done using only gradient information of the m last iterations. Thus the memory requirement will be of $O(mn)$. We will briefly describe this method here (for a detailed description the reader is refered to Byrd *et al.* (1993), Liu and Nocedal (1989) and Nocedal (1980)).

The classical Quasi-Newton method approximates problem (B.1), at every iteration, by a quadratic model

$$m_k(\mathbf{x}) = f(\mathbf{x}_k) + g_k^T(\mathbf{x} - \mathbf{x}_k) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_k)^T B_k(\mathbf{x} - \mathbf{x}_k) \tag{B.2}$$

where $\mathbf{x}_k$ is the current estimate, g_k is the gradient and B_k is a symmetric approximation of the hessian. At each iteration, the method finds the minimum of (B.2), by means of

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \lambda_k H_k g_k \tag{B.3}$$

where H_k is the inverse of the hessian approximation updated at every iteration using the formula

$$H_{k+1} = V_k^T H_k V_k + \rho_k s_k s_k^T \tag{B.4}$$

where

$$\rho_k = \frac{1}{y_k^T s_k}, \qquad V_k = I - \rho_k y_k s_k^T \tag{B.5}$$

and

$$s_k = \mathbf{x}_{k+1} - \mathbf{x}_k, \qquad y_k = g_{k+1} - g_k. \tag{B.6}$$

The matrix H_{k+1} is obtained by updating H_k using the pair $\{s_k, y_k\}$.

The Limited Memory BFGS method, used in this work, has the same framework as presented above, but differs in the matrix H update. Instead of storing the whole matrix H_k, one stores only the most recent m pairs $\{s_i, y_i\}$. Instead of forming H_k explicitly, one forms the matrix-vector product $H_k g_k$ by performing a sequence of inner products involving g_k and the m most recent pairs $\{s_i, y_i\}$.

To carry out the updating process of these approximations to the inverse of the hessian H_k, one has to choose a basic matrix $H_k^{(0)}$ (normally a multiple of a diagonal matrix) and update it m times using BFGS formula (B.4) with the m pairs $\{s_i, y_i\}, i = k-m, ..., k-1$. There is a recursive formula in Nocedal (1980) that takes advantage of the symmetry of the updating formula and it allows to compute the products $H_k g_k$ efficiently. The choice of the initial matrix $H_k^{(0)}$ is crucial, see Gilbert and Lemaréchal (1989), and we actually take $\delta_k I$, where

$$\delta_k = \frac{s_k^T v_k}{y_k^T v_k} \quad \text{with } v_k = \frac{y_k}{\|y_k\|} + \frac{s_k}{\|s_k\|}. \tag{B.7}$$

δ_k serves as a scaling factor that helps to produce the ratio $\frac{\text{function evaluations}}{\text{minimization iterations}}$ close to one.

This basic algorithm can be modified to deal with bounds (Byrd *et al.*, 1993). We describe now how it is done.

To deal with bounds a two step procedure is neccessary. On the first step a generalized Cauchy point $\mathbf{x}^c$ is found. This means moving along the steepest descent direction seeking for the stationary point of a univariate quadratic on the intervals defined by the different steplengths needed to hit the bounds along that direction. So it moves along the piecewise linear path

$$\mathbf{x}(t) = \mathbf{x}_0 + d(t) \tag{B.8}$$

where $\mathbf{x}_0 = \mathbf{x}_k$ and $d(t)$ is the direction

$$d_i(t) = \begin{cases} -tg_i & \text{if } t \leq t_{ij} \\ -t_{ij}g_i & \text{otherwise} \end{cases} \tag{B.9}$$

The t_{ij} are the steplengths needed for each variable i to hit its boundary. This search for a Cauchy point $\mathbf{x}^c$ costs $2mn + O(m^2)n_{\text{int}}$ where n_{int} is the number of intervals.

Once the point $\mathbf{x}^c$ has been found, the $n - s$ variables that are active at that point will remain fix, and the second step of the algorithm seeks for the minimum of quadratic (B.2) on the remaining s free variables, starting from $\mathbf{x}^c$. To accomplish this goal, it generates a Quasi-Newton direction, as in (B.3), with the hessian approximated with the limited memory information, and backtracks if neccessary to find the minimum. In order to make this computation efficiently, a careful matrix manipulation is needed. For further reference see Byrd *et al.* (1993) and Nocedal (1980).

Unité de Recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

Unité de Recherche INRIA Lorraine Technopôle de Nancy-Brabois - Campus Scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 VILLERS LES NANCY Cedex (France)
Unité de Recherche INRIA Rennes IRISA, Campus Universitaire de Beaulieu 35042 RENNES Cedex (France)
Unité de Recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 GRENOBLE Cedex (France)
Unité de Recherche INRIA Sophia Antipolis 2004, route des Lucioles - B.P. 93 - 06902 SOPHIA ANTIPOLIS Cedex (France)

EDITEUR
INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

ISSN 0249 - 6399